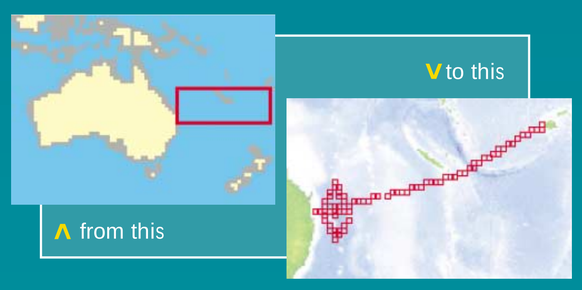
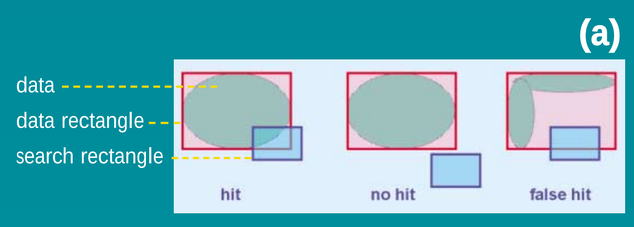
C-squares – a new method for representing, querying, displaying and exchanging dataset spatial extents.

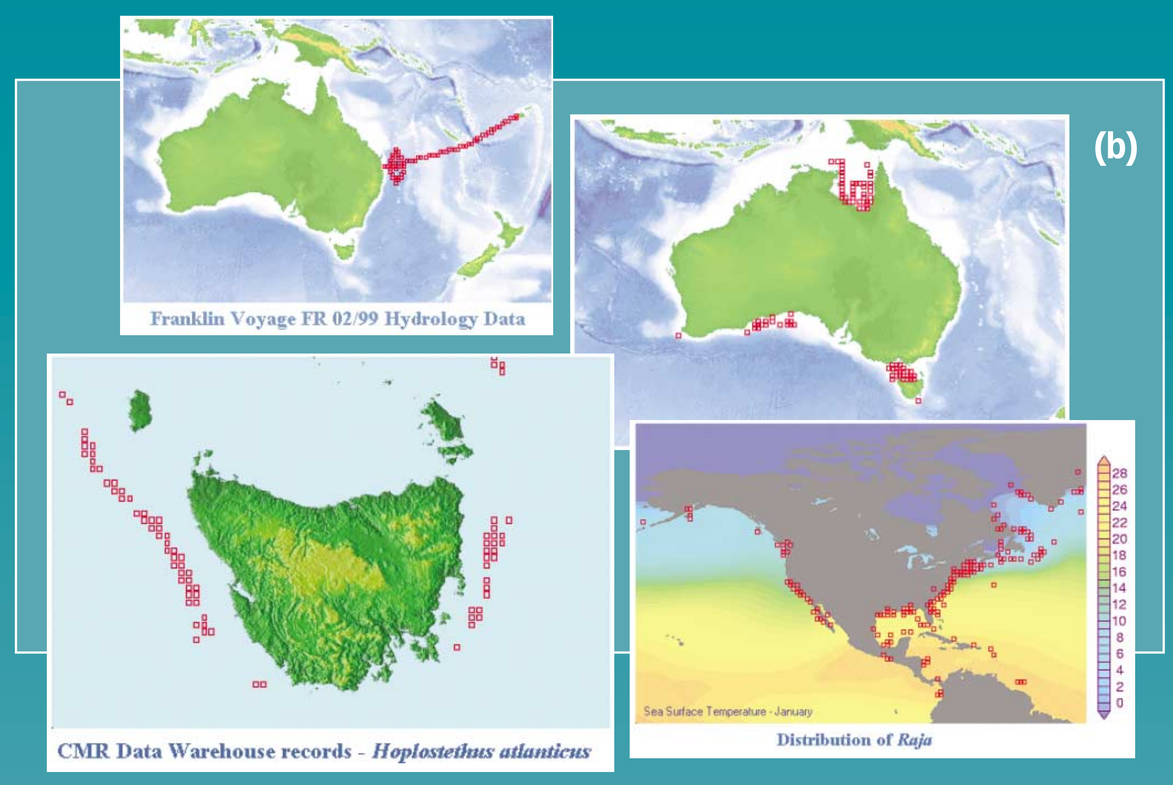
Tony Rees, CSIRO Marine Research, Australia

Conventionally, spatial dataset extents are represented in metadata catalogues (data directories) by bounding rectangles indicating the northern-, southern-, western- and eastern-most limits of the data (FGDC, 1994). Spatial searching of such catalogues can then be supported by allowing the entry of a "search rectangle" which can be used to test for overlaps with equivalent "data rectangles" using simple arithmetic ("greater than / less than" tests) to discover datasets from the region of interest.





While the "overlapping rectangles" test is sufficient to eliminate many datasets whose bounding rectangles do not overlap the search region, it fails to discriminate between true "hits" (those whose bounding rectangle AND actual data extent overlaps the desired search region), and "false hits" (those whose bounding rectangle indicates a match but where, on closer inspection, there is in fact no data from the desired region). This is because, in all cases except where the dataset actually is rectangular in shape, small – or possibly large – portions of the "data rectangle" will be empty. This is easily illustrated with the following examples in theory (a) and in practice (b):



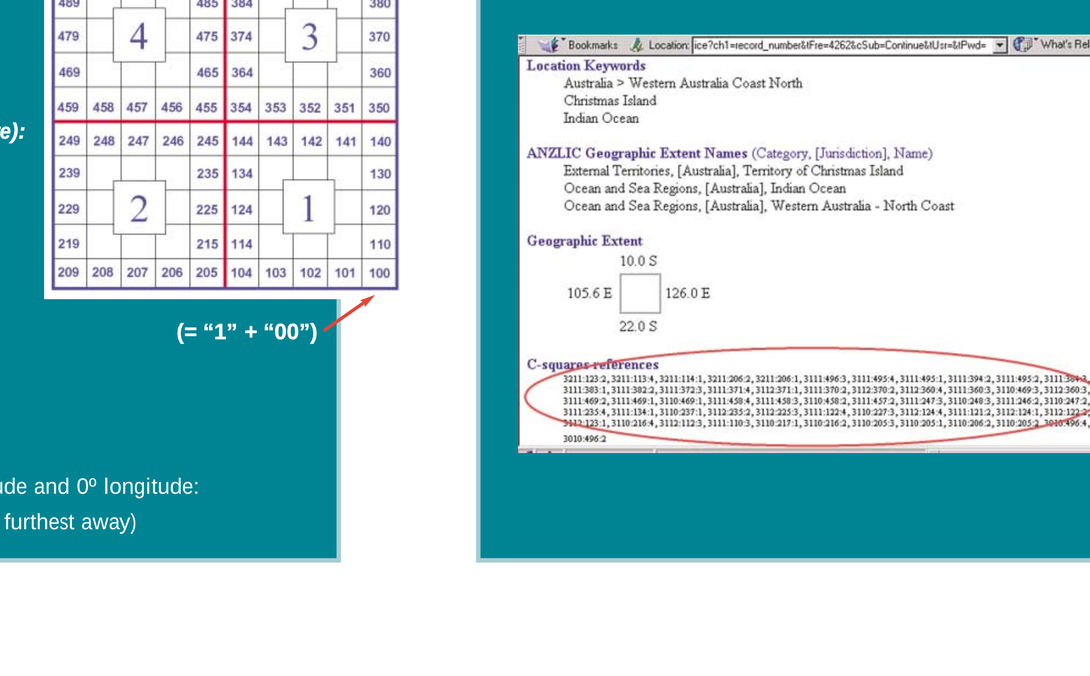
The "c-squares" system addresses this problem by allowing a system or a user to represent any shape or size of dataset spatial extent using a set of numbered squares at any chosen resolution in degrees of latitude and longitude, e.g. 1 x 1º (around 100 x 70 km in temperate latitudes), 0.5 x 0.5º (50 x 35 km), 0.1 x 0.1º (10 x 7 km), or as fine a resolution as may be required. Each square has a pre-defined code based on a recursive subdivision of WMO (World Meteorological Organisation) 10-degree square numbers, e.g. 3013:497 (1 x 1º square) or 3013:497:1 (0.5 x 0.5º square) such that a string of such codes can be written e.g.

3013:497 | 3111:468 | 3111:478 | 3111:479 | 3111:488 | 3111:489 | 3111:499 | 3112:122 | 3112:123 | 3112:131 | 3112:132 | 3112:134 (etc.)

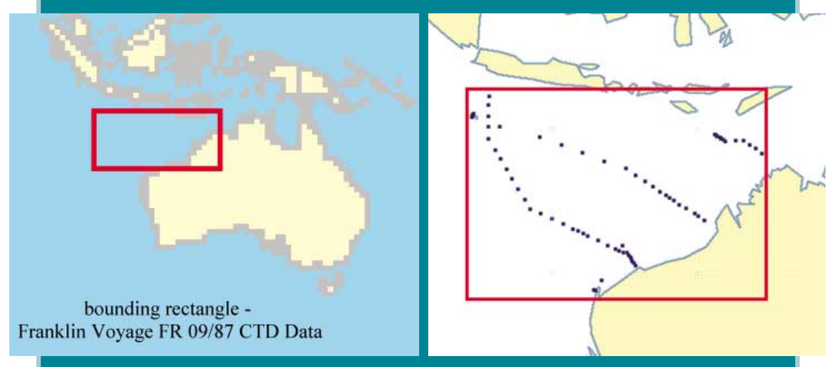
This string can be used as a basis for spatial queries if the query itself is translated to one or more c-squares: for example searching for a c-square 3013:497 will produce a match with the above string, searching for 3013:496 (or 3012) will not. It can also be sent, via the web, to a custom c-squares mapper to produce maps similar to those displayed here, plus used as a portable (data-independent) metadata element representing the dataset spatial extent.

The c-square codes themselves are hierarchical, thus 3013 (10º square) contains 3013:4 (5º square) which contains 3013:497 (1º square) which contains 3013:497:1 (0.5º square), and so on.

The "c-squares" system is particularly useful for representation of oceanographic datasets, which frequently sample only portions of the available ocean surface. However, is equally applicable to terrestrial data – especially for data which has an irregular outline (such as most countries, states or other administrative areas, or natural features), and/or where coverage is patchy or incomplete within the designated boundary.

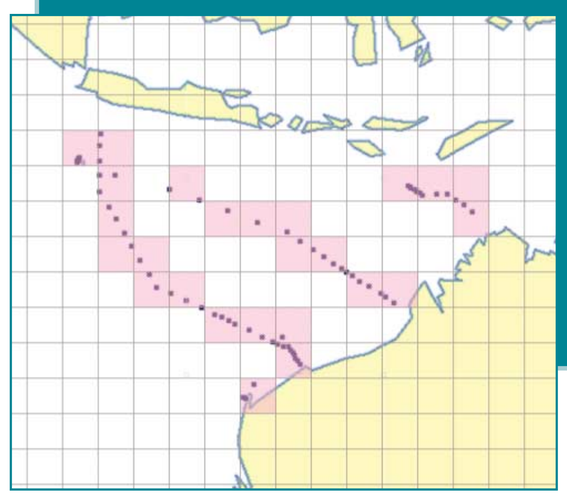


AUTHOR'S AGENCY DATA (TYPICAL):



actual sampling points
(= true dataset extent)

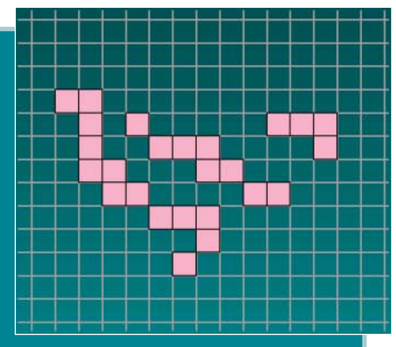
"C-squares" APPROACH:



gridded representation

...gives flexibility to represent a variety of dataset shapes, also patchiness (gaps in data coverage)

HIGHLIGHTED SQUARES:



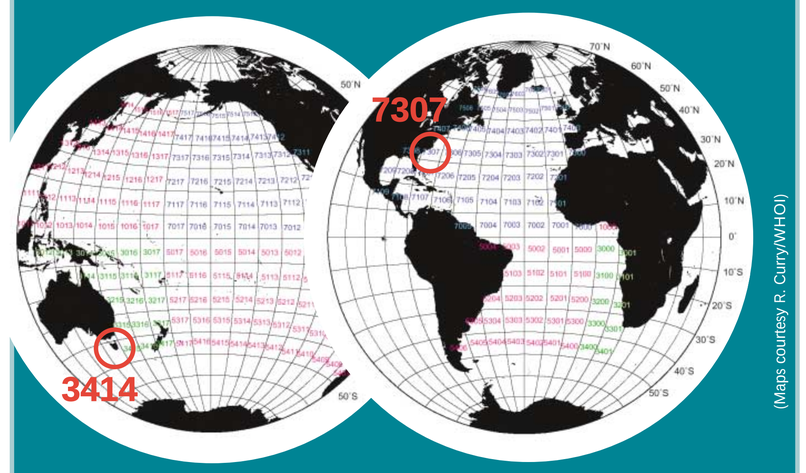
... can be expressed as a set of codes (labels) in an ASCII string, e.g.:

code1 | code2 | code5 | code7 | code13 | code14 | code15 | code21 | (etc.)

- List of codes is potentially more succinct (concise) than original data ...
  - codes potentially terse in themselves
  - multiple points in single square only coded once
  - empty cells not coded

Now has capability for increased precision of querying (on individual square, not bounding rectangle)

WMO 10x10 DEGREE SQUARES IN PRACTICE (EXAMPLES)



(Maps courtesy R. Curry/WHOI)

Washington, DC is in square 7307 (NW global quadrant)
Hobart, Tasmania is in square 3414 (SE global quadrant)

BASIS FOR RECURSIVE SUBDIVISION (e.g. in NW global quadrant)

(Principle as used in Australian "Blue Pages" metadata system, 1996)

10 x 10 deg. square - e.g. 7307

- ***divided as follows ("Blue Pages" nomenclature):***
  - **7307:4** (5 x 5 deg. square)
  - **7307:487** (1 x 1 deg. square)
- ***C-squares then extends this principle recursively, e.g. ...***
  - **7307:487:3** (0.5 x 0.5 deg. square)
  - **7307:487:393** (0.1 x 0.1 deg. square)
  - etc.

(NB, arrangement is mirror image across 0º latitude and 0º longitude: 100 is always closest to the global origin, 499 is furthest away)

C-squares AS EXPLICIT SPATIAL EXTENT CODE/S

C-squares can also be quoted explicitly in metadata records, or any other web document referring to a point or region:

...CAN THEN UTILIZE CAPABILITIES OF A STANDARD INTERNET SEARCH ENGINE, e.g.

For more information on the c-squares system, please visit the c-squares website at: http://www.marine.csiro.au/csquares/ or contact Tony Rees at CSIRO Marine Research, Australia (Tony.Rees@csiro.au).

c-squares

CSIRO MARINE RESEARCH